Development of the BioAssay Research Database (BARD): A User-Friendly Perspective Based on Active Participation from Biologists and Chemists

Eric S. Dawson, Ph.D.

Vanderbilt University

**Integrative Chemogenomics Knowledge Mining
Using NIH Open Access Resources
September 09, 2013**

BARD Direct Contributors

NIH Molecular Libraries – Glenn McFadden, Ajay Pillai

NIH Chemical Genomics Center – *Chris Austin (PI)*, John Braisted, Marc Ferrer, Rajarshi Guha, Ajit Jadhav, Dac-Trung Nguyen, Tyler Peryea, Noel Southall, Henrike Veith

Broad Institute – Benjamin Alexander, Jacob Asiedu, Kay Aubrey, Joshua Bittker, Steve Brudz, Simon Chatwin, Paul Clemons, Vlado Dancik, Siva Dandapani, Andrea DeSouza, Dan Durkin, David Lahr, Jeri Levine, Judy McGloughlin, Phil Montgomery, Jose Perez, *Stuart Schreiber (PI)*, Gil Walzer, Xiaorong Xiang

University of New Mexico – Cristian Bologa, Steve Mathias, Tudor Oprea, Larry Sklar, Oleg Ursu, Anna Waller, Jeremy Yang

University of Miami – Saminda Abeyruwan, Hande Küküc, Vance Lemmon, Ahsan Mir, Magdalena Przydzial, Kunie Sakurai, Stephan Schürer, Uma Vempati, Ubbo Visser

Vanderbilt University – Eric Dawson, David Weaver, Shaun Stauffer, Sabuj Pattanayek, Craig Lindsley
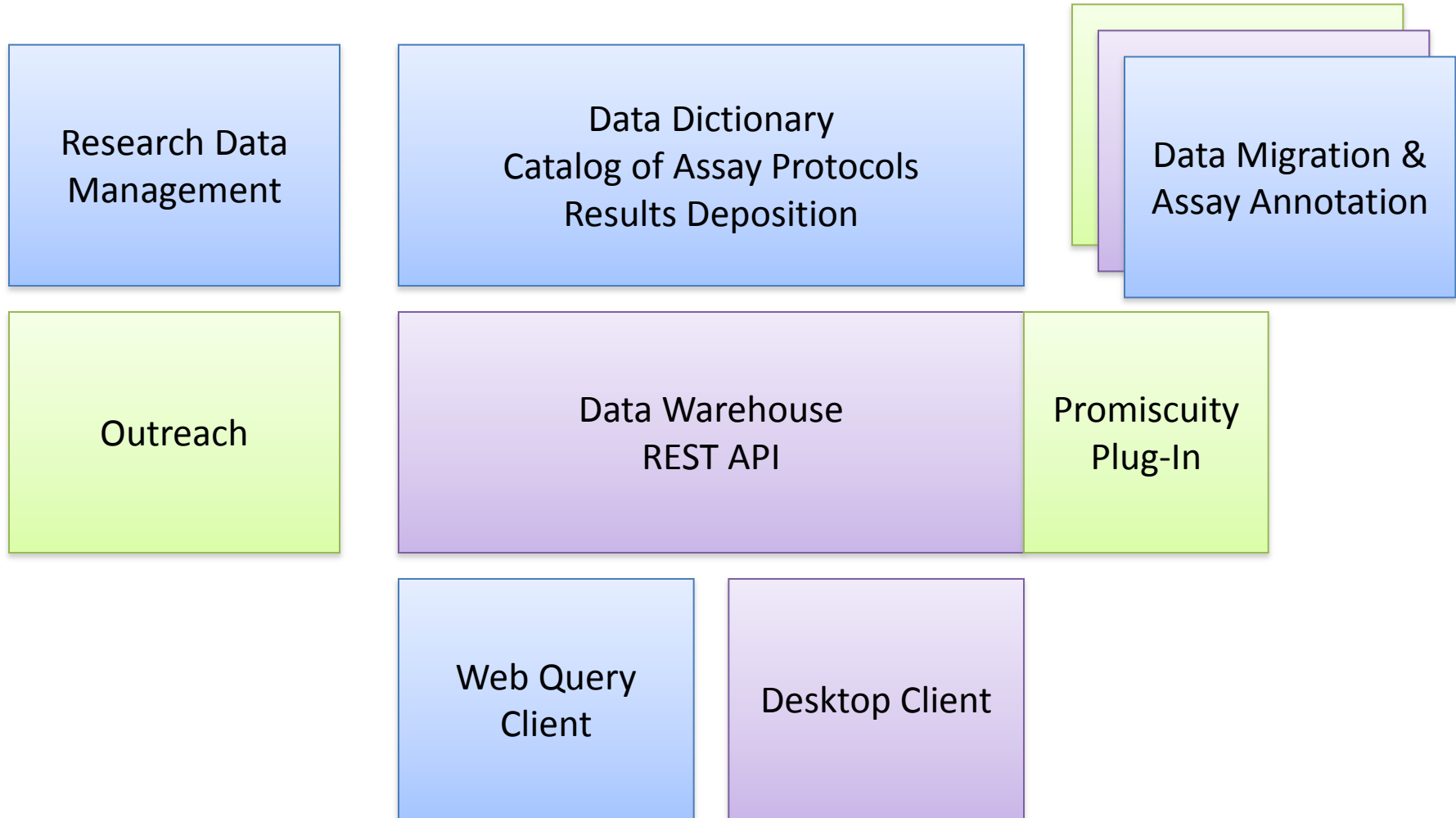
Sanford-Burnham Medical Research Institute – "T.C." Chung, Jena Diwan, Michael Hedrick, Gavin Magnuson, Siobhan Malany, Ian Pass, Anthony Pinkerton, Derek Stonich
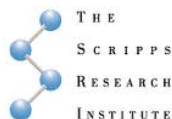
**Scripps Research Institute** – Yasel Cruz, Mark Southern

BARD Component Development

Research Data Management	Data Dictionary Catalog of Assay Protocols Results Deposition	Data Migration & Assay Annotation
Outreach	Data Warehouse REST API	Promiscuity Plug-In
	Web Query Client	Desktop Client

- ■ Broad, Sanford-Burnham, Scripps, U of Miami
- ■ NCGC
- ■ UNM & Vanderbilt

BARD
BioAssay Research Database

BARD Technological Components

Operational & Analytical data stores with public APIs to enable community development
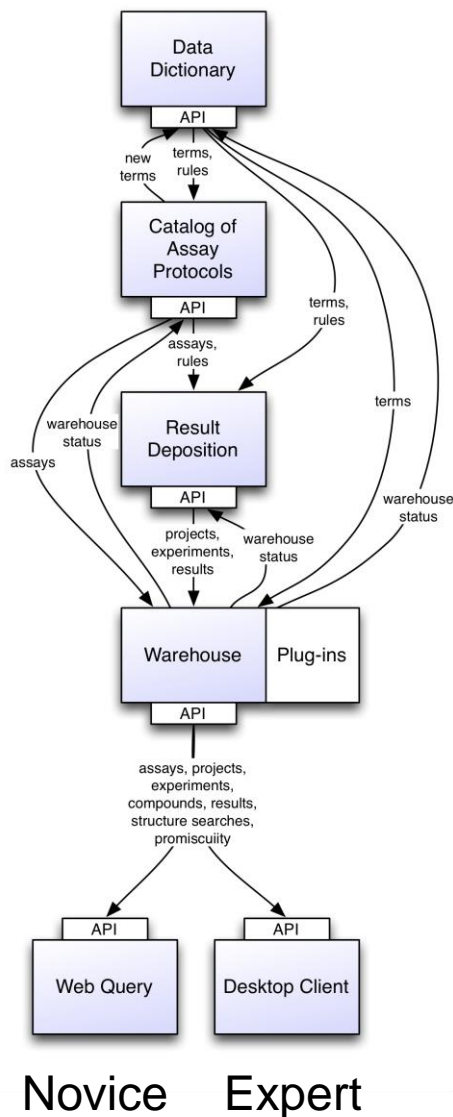
Define & Register Assays
Data Dictionary – std terms
Catalog of Assay Protocols

High Quality Data & Result Deposition
Calculations & Results
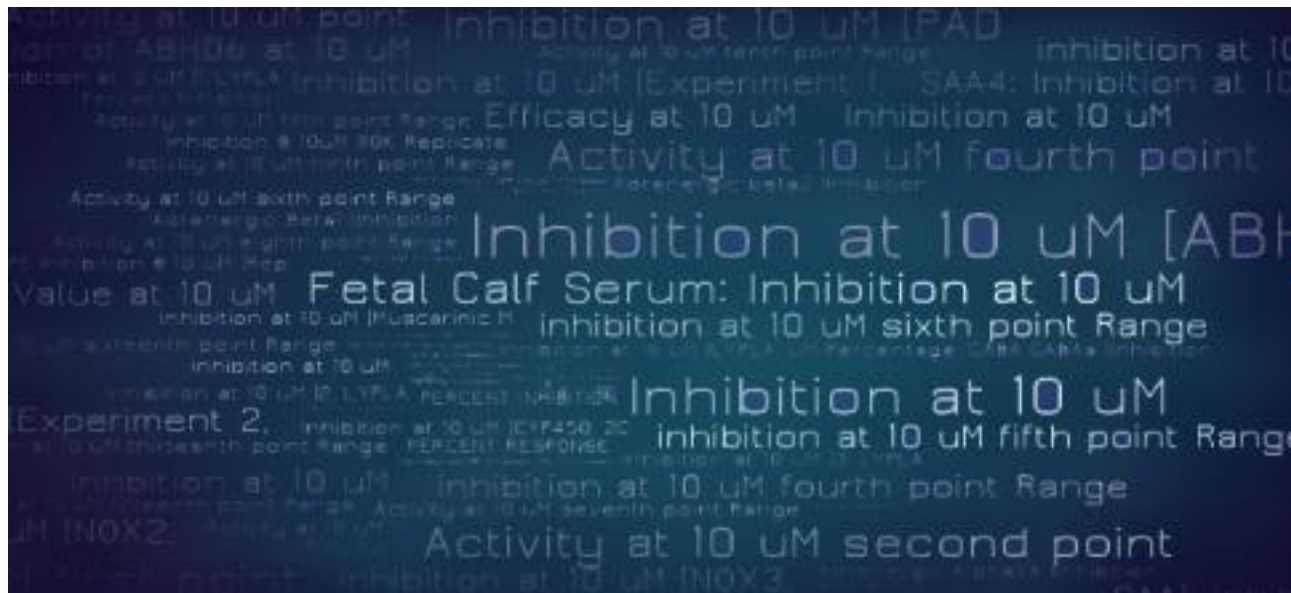Project-experiment association

Query & Interpret Information
Intuitive Guided Queries
Cross Assay & SAR centric views
Advance applications
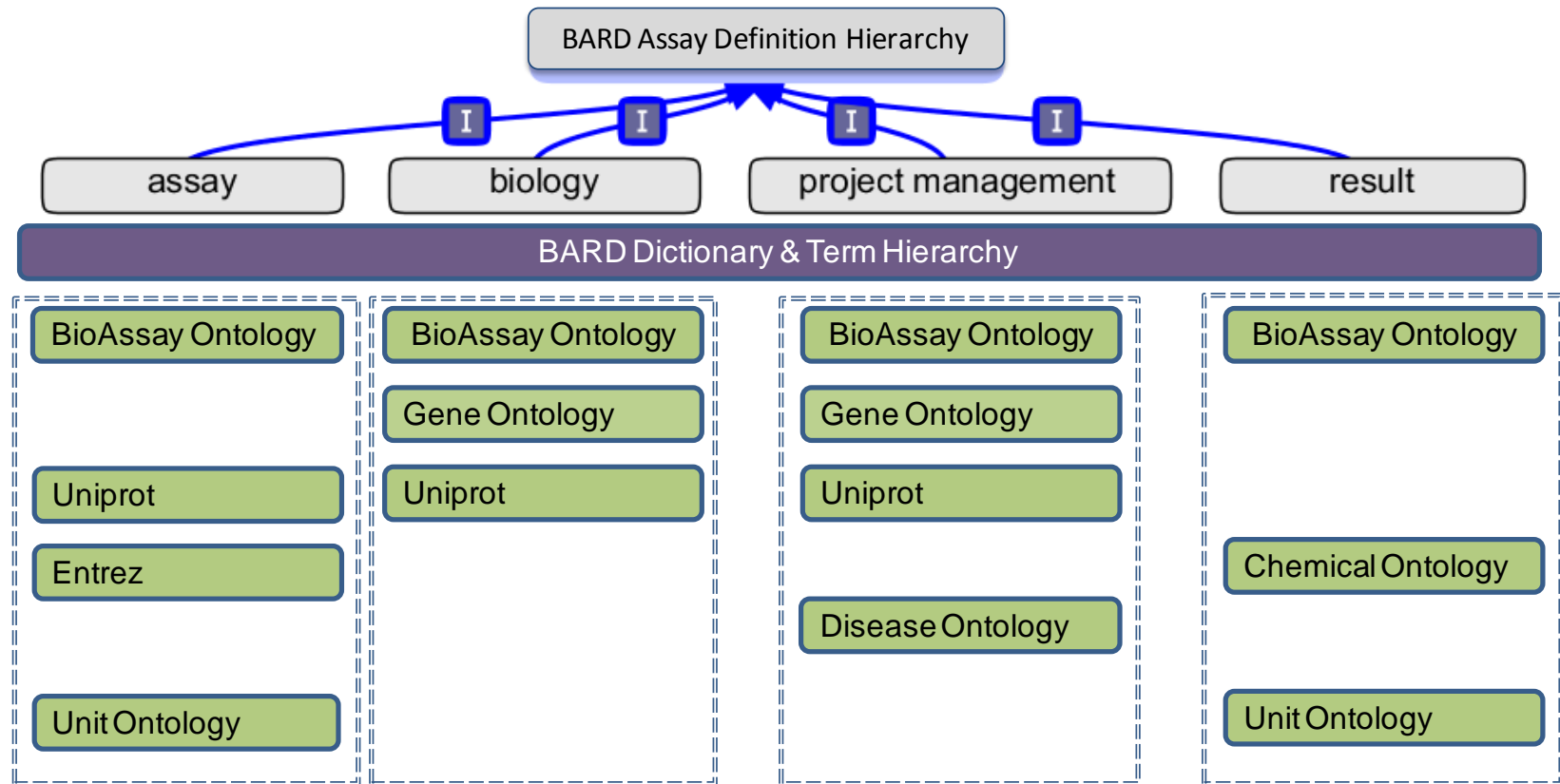
Enable Hypothesis Generation

Novice   Expert

Data Dictionary Development

The Power of a Common Language: Intuitive Scientific Terms from Engaging Chemists and Biologists

# RDM: BARD data dictionary & associated ontologies

Principles
- Primary purpose is to define assays in a structured way
- Use language that experimentalists use whenever possible
- Leverage the expertise of others; don't duplicate qualified existing term authorities
- Core terms are concepts that change infrequently
- Dictionaries are referenced by core for ease of modification

BARD
BioAssay Research Database

Catalyst Meeting Sessions (Broad)

- **Research Data Management**
  - Included stakeholders (chemists / biologists)
  - Use of common terms (positive allosteric modulator)
  - Integration with public ontologies (best possible extent)

Fold-Shift = 100

Fold Maximum = 1.0

% of Vehicle Treated Response

110
100
90
80
70
60
50
40
30
20
10
0

-13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1

log [Othosteric Agonist], M

— + Vehicle
— + PAM (Same Max)

Catalyst Meeting Sessions (Broad)
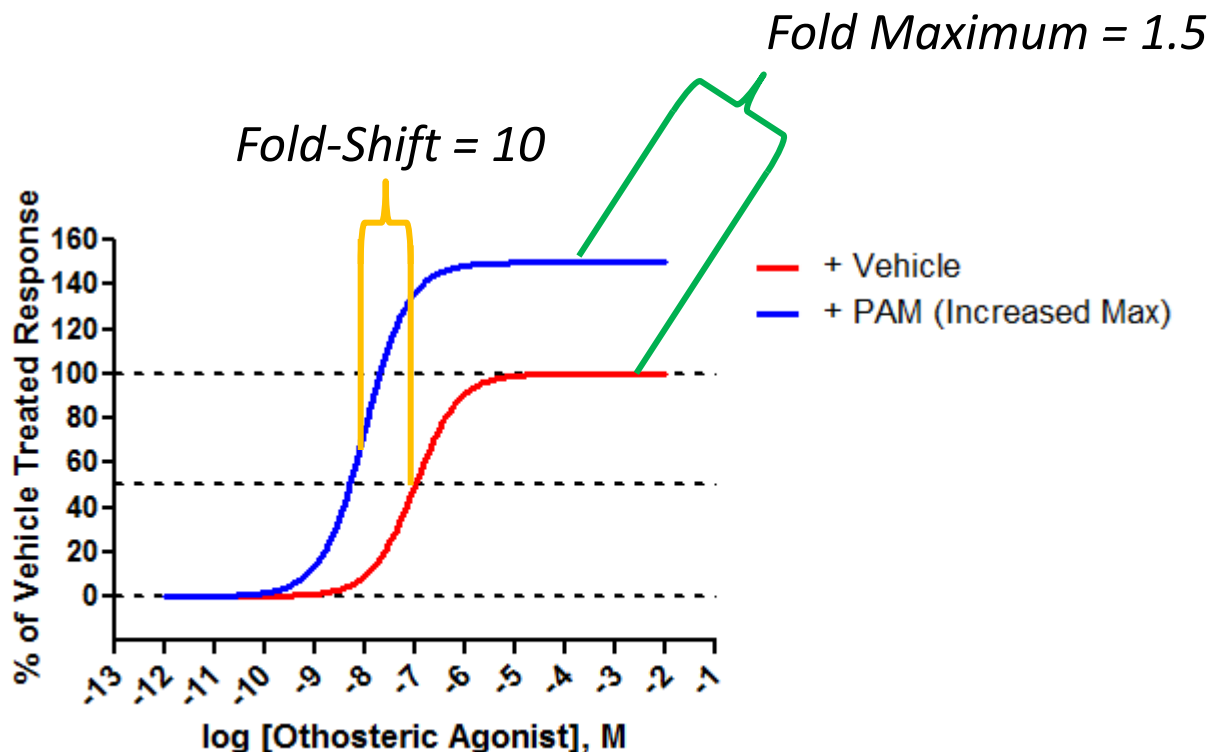
- **Research Data Management**
  - Included stakeholders (chemists / biologists)
  - Use of common terms (positive allosteric modulator)
  - Integration with public ontologies (best possible extent)

Fold Maximum = 1.5

Fold-Shift = 10
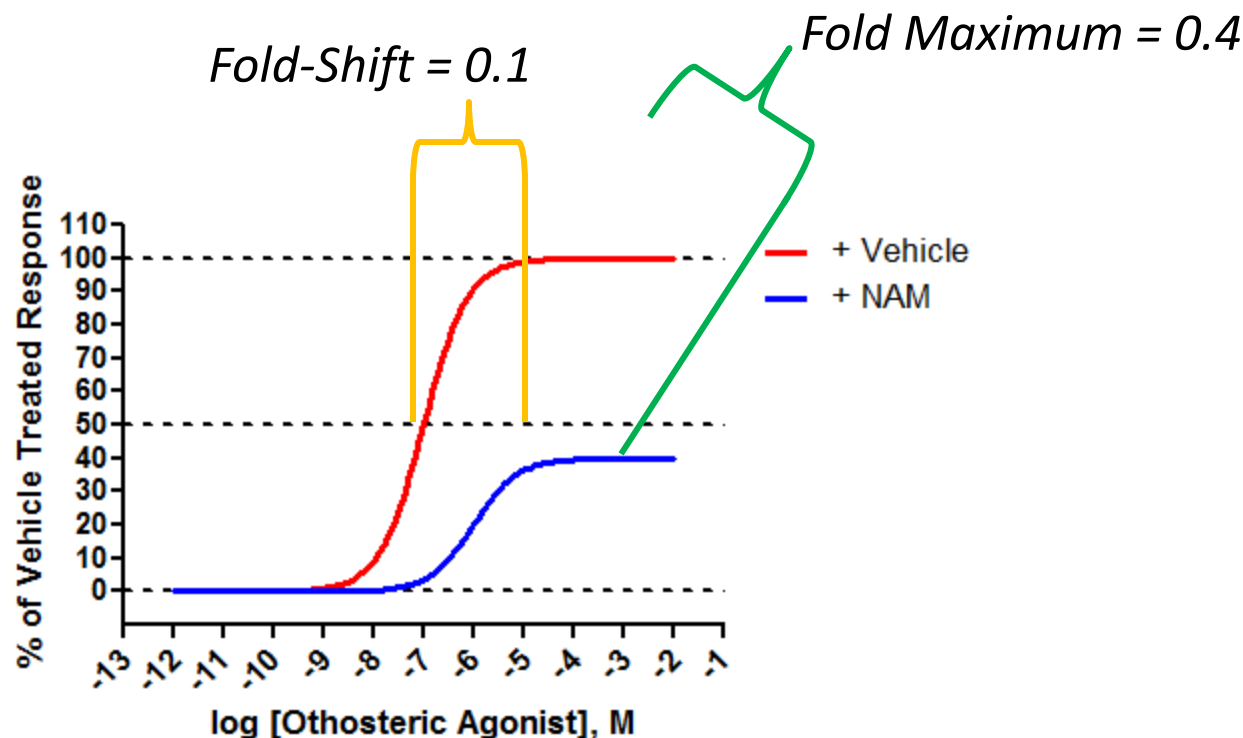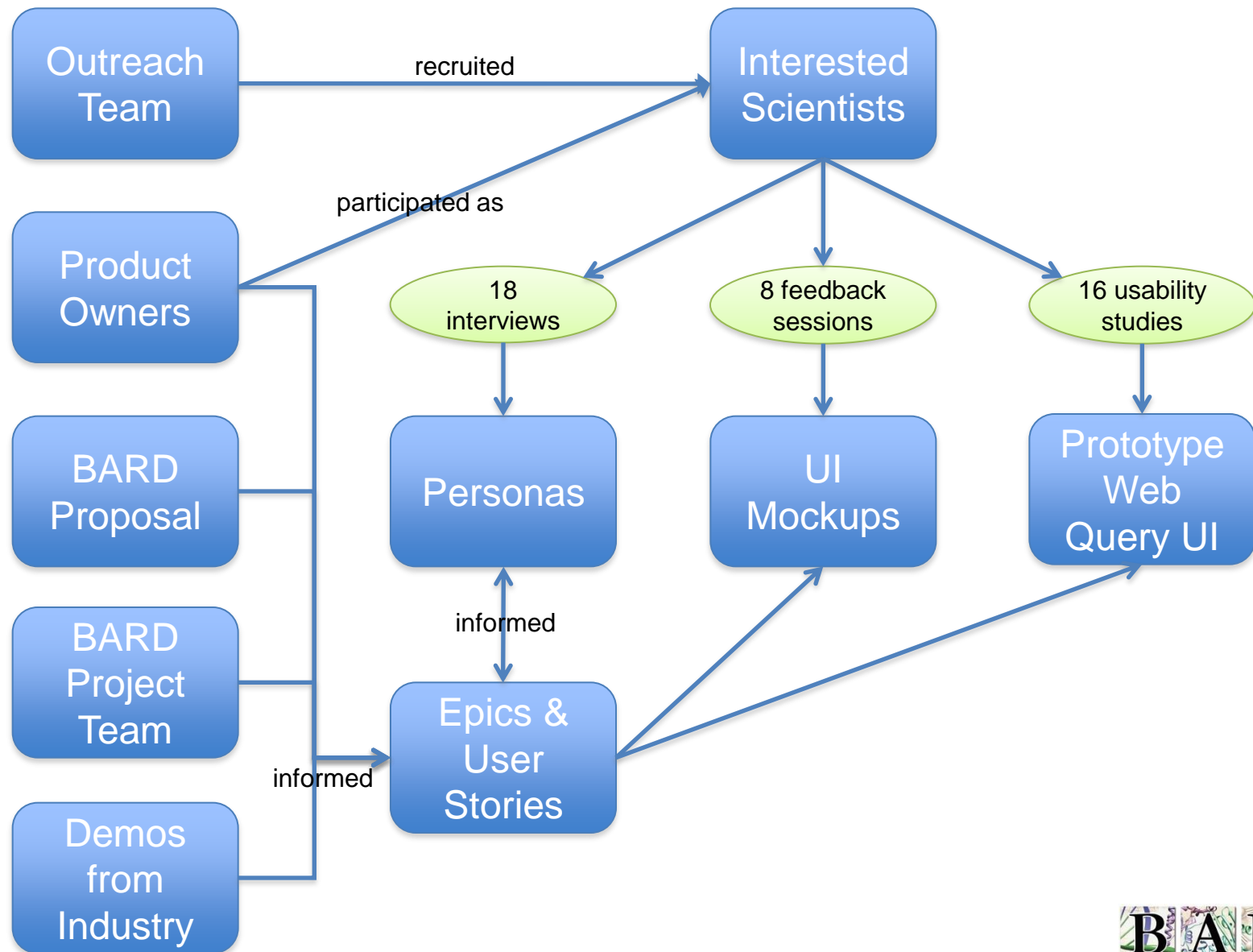
+ Vehicle
+ PAM (Increased Max)
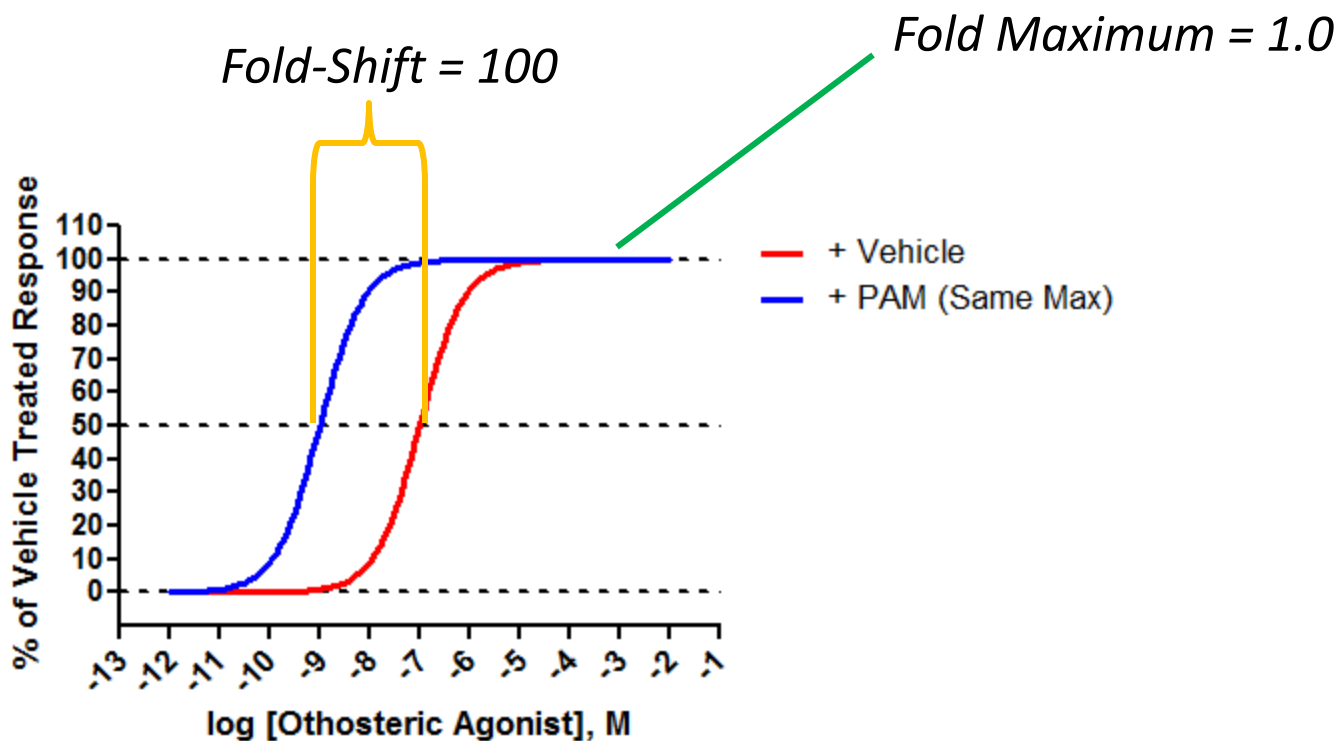
Catalyst Meeting Sessions (Broad)

- **Research Data Management**
  - Included stakeholders (chemists / biologists)
  - Use of common terms (negative allosteric modulator)
  - Integration with public ontologies (best possible extent)

Fold-Shift = 0.1

Fold Maximum = 0.4

+ Vehicle

+ NAM

% of Vehicle Treated Response

log [Othosteric Agonist], M

Iterative Usability Studies (Broad)

Iterative Usability Studies (Broad)

- **User Interface Demonstrations (web client)**
  - Included stakeholders (chemists / biologists)
  - Guided tours plus user-driven interface exploration
  - Suggested and user-provided query exploration
  - *Data quality limited based on early-stage engagement*

- **User Interface Demonstrations (CAP)**
  - Included stakeholders (HTS biologists / Data management)
  - Guided tours plus user-driven interface exploration
  - Suggested and user-provided annotation exploration
  - *Project level data quality limited based on early-stage*

Iterative Usability Results (Broad)

- **Research Data Management**
  - Assay Definition Standards
  - Result Aggregation
  - Integration with public ontologies

- **Good System Architecture**
  - 'Google-like' User Interfaces
  - Auto suggest search queries by annotations
  - Simultaneous search of compounds, assays and projects
  - Structure based searches
  - Filtering of search results by annotations / metadata
  - Detailed views of data at each level (Compound, Assay / Experiment, Project)
  - Visualization via a molecular spreadsheet
  - Export from spreadsheet to common formats (Excel, CSV, PDF)
  - Mobile client app with basic search functionality
  - Integration infrastructure to deploy public algorithms
  - Iterative development cycle informed by regular user feedback

- **Next-generation cheminformatics analysis**
  - Requires *high quality data* and good system architecture

BARD
BioAssay Research Database

Outreach Efforts

Scientific Society Targeted Meetings for BARD Awareness

OPEN PHACTs Meeting: Tudor Oprea (Univ. New Mexico)
(Vienna - August 2012)

Academic Assay and Screening Workshop (ASW)
(Baltimore, MD 09-20/22-2012)

SLAS Presentation: Thomas Chung (Sanford Burnham)
(Orlando, FL 01/12-01/16/2013)
Afternoon SIG Session (BARD mini-symposium)
Noel Southall (NCATS) / Andrew Napper (Meeting Chair, SLAS)

ACS Invited talk - CINF section
(Spring 2013 – New Orleans)
Raj Guha (NCATS); Jun (Luke) Haun, Kansas – session chair

BioAssay Research Database

Outreach Efforts

BARD Collaborators coordinated publication strategy
(e.g. assay de-duplication, app use-case)

Early engagement of Biologists in University settings
(Fall 2013 – following BARD public beta launch)
Targeting Washington University (Joshua Swamidass), UNC (Alex Tropsha); St. Jude's (Kip Guy), VU biologists (HTS core)

BARD local servers

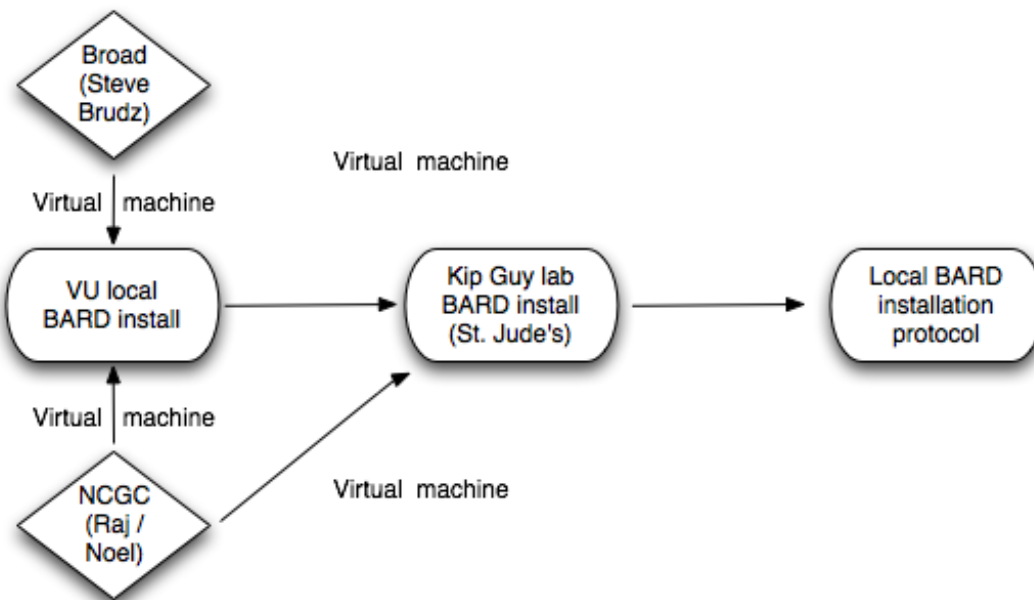## Initial BARD local server installations scheduled
(Vanderbilt University, St. Jude's – Kip Guy lab)

Local BARD Instance

Staged Install plan (VM first)

Year 2 deployment of native hardware server solution for performance testing (VU)

Successful installation of external 'private' BARD instance supports generation of novel IP
(academic / industrial partnerships)

BARD Contributors & Stakeholders

Direct Contributors – Current MLPCN Centers *(generators of > 85% of PubChem data)*

Stakeholders & Advisors: Architecture Advisory (Wash U)**, Technical Advisory Group** (NIH, EBI, NIBR,Takeda)**, Requirements & Usability Feedback Group**

* **MLPCN Center**

Development of the BioAssay Research Database (BARD): A User-Friendly Perspective Based on Active Participation from Biologists and Chemists

Eric S. Dawson, Ph.D.

Vanderbilt University

eric.dawson@vanderbilt.edu

**Integrative Chemogenomics Knowledge Mining
Using NIH Open Access Resources
September 09, 2013**